Selection of Bradyrhizobium or Ensifer symbionts by the native Indian caesalpinioid legume Chamaecrista pumila depends on soil pH and other edaphic and climatic factors

Rathi, Sonam; Tak, Nisha; Bissa, Garima; Chouhan, Bhawana; Ojha, Archana; Adhikari, Dibyendu

Published in:
FEMS Microbiology Ecology

DOI:
10.1093/femsec/fiy180

Publication date:
2018

Document Version
Peer reviewed version

Selection of *Bradyrhizobium* or *Ensifer* symbionts by the native Indian caesalpinioid legume *Chamaecrista pumila* depends on soil pH and other edaphic and climatic factors

Sonam Rathi[1], Nisha Tak[1*], Garima Bissa[1], Bhawana Chouhan[1], Archana Ojha[2], Dibyendu Adhikari[2], Saroj K. Barik[2], Rama Rao Satyawada[3], Janet S. Sprent[4], Euan K. James[5] and Hukam S. Gehlot[1]

[1]BNF and Microbial Genomics Lab., Department of Botany, Center of Advanced Study, Jai Narain Vyas University, Jodhpur- 342001, Rajasthan, India.

[2]Department of Botany, North-Eastern Hill University, Shillong-793022, Meghalaya, India.

[3]Department of Biotechnology and Bioinformatics, North-Eastern Hill University, Shillong-793022, Meghalaya, India.

[4]Division of Plant Sciences, University of Dundee at the James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK.

[5]The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK.

[*]Corresponding author- Nisha Tak, BNF and Microbial Genomics Lab., Department of Botany, Center of Advanced Study, Jai Narain Vyas University, Jodhpur- 342001, Rajasthan, India

Tel.: +919461212680, Email: nishatak13@gmail.com

One sentence summary: The novel rhizobial microsymbionts that have co-evolved with the basal legume *Chamaecrista pumila* in multiple niches in India are either *Bradyrhizobium* or *Ensifer* depending on the soil pH and other edaphic/climatic factors.

ABSTRACT

Nodules of *Chamaecrista pumila* growing in several locations in India were sampled for anatomical studies and for characterization of their rhizobial microsymbionts. Regardless of their region of origin, the nodules were indeterminate with their bacteroids contained within symbiosomes which were surrounded by pectin. More than 150 strains were isolated from alkaline soils from the Thar Desert (Rajasthan), wet-acidic soils of Shillong (Meghalaya), and from trap experiments using soils from four other states with different agro-ecological regions. Molecular phylogenetic analysis based on five housekeeping (*rrs*, *recA*, *glnII*, *dnaK*, *atpD*) and two symbiotic (*nodA*, *nifH*) genes was done for selected strains. *Chamaecrista pumila* was shown to be nodulated by niche-specific diverse strains of either *Ensifer* or *Bradyrhizobium* in

alkaline (Thar Desert) to neutral (Tamil Nadu) soils and only *Bradyrhizobium* strains in acidic (Shillong) soils. Concatenated core gene phylogenies showed four novel *Ensifer*-MLSA types and nine *Bradyrhizobium*-MLSA types. Genetically diverse *Ensifer* strains harbored similar *sym* genes which were novel. In contrast, significant symbiotic diversity was observed in the *Bradyrhizobium* strains. The *C. pumila* strains cross-nodulated *Vigna radiata* and some wild papilionoid and mimosoid legumes. It is suggested that soil pH and moisture level played important roles in structuring the *C. pumila* microsymbiont community.

Key words: *Ensifer*, *Bradyrhizobium*, MLSA, Symbiotic genes, fixation threads, JIM5

INTRODUCTION

The Leguminosae (Fabaceae) is the third largest angiosperm family and is of immense ecological and economic importance. It has traditionally been divided taxonomically into three sub-families (Lewis *et al*., 2005), but has recently been reclassified into six subfamilies consisting of the Papilionoideae (largely unchanged) and five new subfamilies created from the old paraphyletic Caesalpinioideae: the Caesalpinioideae (sensu stricto), Duparquetioideae, Cercidoideae, Detarioideae and Dialioideae (LPWG 2017). The ex-Mimosoideae subfamily has been downgraded and is now referred to as the mimosoid clade of the Caesalpinioideae (s.s.). Many legumes are unique in forming symbioses with soil bacteria termed rhizobia which are housed in root- (sometimes stem-) borne structures called nodules, and which fix atmospheric nitrogen into ammonia and then into usable, non-toxic forms for the benefit of the host plant (Sprent *et al.*, 2017). The vast majority of papilionoid legumes nodulate, but most of the genera within the newly-circumscribed subfamilies in the Caesalpinioideae sensu lato, which are regarded as basal in the evolution of legumes, are non-nodulated. Those few genera that do nodulate are confined to the Caesalpinioideae sensu stricto (as well as to the largely nodulated mimosoid clade which is nested within it). The basal position of the nodulated members of the Caesalpinioideae, and the fact that they are scattered amongst the largely non-nodulated former tribes, Cassiae and Caesalpiniae, means that they are pivotal to studies into the origins of the legume-rhizobium symbiosis (Doyle 2011; Sprent *et al.*, 2013, 2017). Indeed, nodules on most caesalpinioid legumes studied to date apparently possess "primitive" features, such as the retention of the rhizobium symbiotic form, the bacteroids, within cell wall-bound "fixation threads" (de Faria *et al.*, 1987; Naisbitt *et al.*, 1992; Fonseca *et al.*, 2012).

The nodulated caesalpinioids described to date consist of eight genera, seven of which are in the former tribe Caesalpiniae, and only one genus, *Chamaecrista*, in the former tribe Cassiae, wherein it sits with its non-nodulated sister genera *Cassia* and *Senna* (Irwin and Barneby 1982; Lewis *et al*., 2005; Sprent *et al*., 2013, 2017). *Chamaecrista* is by far the largest genus of nodulating caesalpinioids, comprising more than 330 species distributed mostly in the New World, wherein 256 species are found only in Brazil, the centre of radiation for the genus (Barneby 1994), including several endemic species, as well as some that are considered as invasive (dos Santos *et al*., 2017). Most *Chamaecrista* species are shrubs or small herbs, but some are trees. Some species are pantropical, and one of these, *C. rotundifolia*, is widely used as a forage legume in Australia, where it is known as "Wynn Cassia" (Lafay and Burdon 2007).

All *Chamaecrista* species so far studied are nodulated, and nodulation appears to be a generic trait (Naisbitt *et al*., 1992; Sprent 2009; Parker and Rousteau 2014; Beukes *et al*., 2016; dos Santos *et al*., 2017). There are relatively few reports on the rhizobial microsymbionts associated with caesalpinioid legumes, including *Chamaecrista*, and most reports are from the New World. Rhizobia are now considered to be a polyphyletic group of bacteria scattered amongst non-nodulating genera in the Alpha- and Betaproteobacteria (Gyaneshwar *et al*., 2011; Peix *et al*., 2015; Andrews and Andrews 2017), but the most numerous and widely dispersed rhizobial genus in both host range and geographical terms is *Bradyrhizobium*, which is also considered to be the most likely candidate for being the ancestral rhizobial symbionts (Menna *et al*., 2009; Parker 2015; Sprent *et al*., 2017). Almost all rhizobia so far isolated from caesalpinioids have been described as belonging to *Bradyrhizobium* (Parker 2000, 2015; Moulin *et al*., 2004; Fonseca *et al*., 2012; Yao *et al*., 2014, 2015), and this is also the case for *Chamaecrista* (Moreira *et al*., 1998; Parker and Kennedy 2006; Parker 2012; Beukes *et al*., 2016), except for one report of *Mesorhizobium* associated with *C. ensiformis* a tree native to Brazil, although the nodulation ability of this strain has not yet been confirmed (Moreira *et al*., 1998). The preference of *Chamaecrista* for *Bradyrhizobium* in the neotropics was recently confirmed by dos Santos *et al*. (2017) in a detailed study of symbionts of herbaceous *Chamaecrista* species in the north east Brazilian state of Bahia, which has several endemic species (Lewis 1987).

There are few reports on nodulation of *Chamaecrista* and its associated microsymbionts from the Old World even though there are several native species. India, for example, contains

two endemic species (*C. kolabensis* and *C. nilgirica*), a few native ones as well as several pantropical, introduced species (Lewis 2005), but no nodule isolates have been characterized from any of them so far. Gehlot *et al*. (2012) in their study of legumes native to the Thar Desert and other arid regions of Western Rajasthan (RJ), reported nodulation of *Chamaecrista pumila*, a widespread native species in the Indian subcontinent and South East Asia (Lewis 2005). In the present investigation the rhizobia associated with *C. pumila* naturally growing in two contrasting eco-climatic regions within India, the semi-arid Thar Desert of Rajasthan (RJ) and the wet and humid sub-Himalayan region of Shillong in Meghalaya (ME) were isolated. Trapping experiments using soil from various sites across India, including Shillong (ME), the Thar Desert (RJ) and the states of Gujarat (GJ), Tamil Nadu (TN), Jharkhand (JH) and Uttarakhand (UT) were also performed. The aims of the study were to (1) catalogue nodulation of *C. pumila* in contrasting eco-climatic regions in India and to describe the structure and ultrastructure of the nodules using light and electron microscopy; (2) to identify and characterize the diversity of the native rhizobial microsymbionts and to compare them with those isolated from *Chamaecrista* species in the New World, especially Brazil, the centre of radiation; and (3) determine if the environment (soil, climate) affects the selection of symbionts by this widespread and ecologically important legume.

MATERIALS AND METHODS

Nodule sampling, rhizobial trapping and isolation

Chamaecrista pumila plants were sampled in the field in RJ (Fig. S1A–B) and ME (Fig. S1C–D), germplasm was collected for trap experiments (Fig. S1E), and nodules (Fig. S1F–G) were sampled when present. For rhizobial trapping experiments soil was collected from the rhizosphere of *C. pumila* from alkaline sites in the Thar Desert in RJ and from acidic sites from Shillong in ME. Soil for trapping was also collected from other Indian states, such as GJ, TN, JH and UT (Fig. 1, S2). The rhizospheric soils collected from the various sampling sites were analyzed for various parameters using standard methods (Gehlot *et al.*, 2012). For some of the sampling sites soil data were available online (Table S1 and S2). To trap rhizobia, surface-sterilized germinated seeds were transferred into plastic pots containing soil of a specific site as described in Sankhla *et al*. (2017). Plants were harvested after 6-8 weeks to check their nodulation status. Pink and healthy nodules of different developmental stages were collected for

external morphological and anatomical studies as well as for isolation of rhizobia. Root nodules were thoroughly surface-sterilized, and after puncturing the nodules, the exudate was streaked onto Congo Red-Yeast Extract Mannitol Agar (CR-YEMA) medium (Somasegaran and Hoben 1994) for isolation and culturing of rhizobia using the method described in Gehlot *et al*. (2012, 2013). Streaked plates were incubated for 2–8 days at 28±2^{o}C in an incubator and checked regularly for growth of rhizobial colonies which are white, raised, concave and opaque with entire margins.

Nodule anatomy

Some healthy nodules from the field and from the trap experiments were also fixed in 1% glutaraldehyde for anatomy studies; these were embedded in resin and sectioned for light and transmission electron microscopy (TEM) as described by Elliott *et al*. (2007). Some sections were immunogold labelled using the monoclonal antibody, JIM5, which recognises a pectin epitope in plant cell walls (Fonseca *et al*., 2012), and with a polyclonal antibody which recognises the NifH protein of nitrogenase (dos Reis Junior *et al*., 2010; Gehlot *et al*., 2013).

Preparation of genomic DNA and amplification of 16S rRNA, housekeeping, and symbiosis-essential genes

Genomic DNA of the rhizobial strains was extracted using the phenol-chloroform method described by Cheng and Jiang (2006). The purified DNA at a concentration of 100–1000 ng µl^{-1} was used as a template for amplification of small subunit ribosomal RNA (16S rRNA), four protein-coding housekeeping genes (*dnaK, atpD, recA* and *glnII*) and two symbiotic genes (*nifH* and *nodA*). The list of primers and the thermal cycling conditions used for amplification and sequencing of various genes is given in Table S3.

PCR amplification of the nearly full-length 16S rRNA gene was performed using two universal primer sets: (i) 18F and 1492R (ii) fD1 and rD1 (Weisburg *et al*., 1991). The reaction mixture was prepared as reported in Sankhla *et al*. (2017) and subjected to the thermal cycling conditions in Table S3. The amplified PCR product (size approx. 1500 bp) along with 500 bp DNA ladder (Genei Bangalore) were electrophoresed in a 0.8% agarose gel prepared in TAE buffer and visualized by ethidium bromide staining on a BIO-RAD Gel Doc system.

For multi locus sequence analysis (MLSA) of selected rhizobial strains, four protein-coding housekeeping genes (*dnaK, atpD, recA* and *glnII*) were amplified and sequenced.

Amplification of the *dnaK* (Stepkowski *et al*., 2003), *atpD* (Martens *et al*., 2008), *recA* (Gaunt *et al*., 2001) and *glnII* (Turner and Young 2000) genes in *Ensifer* strains was achieved using the primer pairs and cycling conditions given in Table S3. For *Bradyrhizobium* strains, the primers described by Stepkowski *et al*. (2003) were used for amplification of *dnaK*, Stepkowski *et al*. (2005) for *recA* and *glnII*, and Gaunt *et al*. (2001) for *atpD*. The PCR reaction mix was the same as used for amplification of the 16S rRNA gene except that the volume of $MgCl_2$ (25 mM) was 2.5 µl for *atpD* and 3.75 µl for *dnaK*. The thermal cycling conditions used for amplification of the respective genes in *Bradyrhizobium* strains are given in Table S3.

Two symbiosis-essential genes (*nifH* and *nodA*) were amplified and sequenced for selected strains. For amplification of a 650 bp fragment of the *nodA* gene in *Ensifer* strains the primers nodA1 and nodA2 (Haukka *et al*., 1998) were used and for PCR amplification of a 550 bp *nodA* gene fragment in *Bradyrhizobium* strains the primers nodAf.brad and nodAr.brad were used (Chaintreuil *et al*., 2001). Amplification of a 750 bp *nifH* fragment in both *Ensifer* and *Bradyrhizobium* strains was performed using the primers described by Laguerre *et al*. (2001). The PCR mix for amplification of various *sym* genes was prepared as reported in Sankhla *et al.* (2017) and the thermal cycling conditions are given in Table S3.

DNA sequencing and phylogenetic analysis

The amplified PCR products of various genes (60 µl containing approx.100 ng μl^{-1} DNA) were sent to SciGenom Labs Private Ltd. Cochin, India for exo-sap purification and Sanger sequencing using the Applied Biosystems (ABI) platform. All the raw sequences were analyzed, edited and assembled to nearly full length using the Gene Tool Lite software (www.biotools.com). Basic Local Alignment Search Tool (BLAST) was used for sequence similarity searches in the National Centre for Biotechnology Information (NCBI) nucleotide sequence database (http://blast.ncbi.nlm.nih.gov) (Altschul *et al*., 1990). The complete analyzed sequences of the genes were submitted to the NCBI GenBank database using the Sequin software, and accession numbers were obtained.

Phylogenetic analyses were carried out using the sequences of type/reference strains and close relatives retrieved from the NCBI GenBank database with the help of MEGA version 7.0 (Kumar *et al*., 2016). Multiple sequence alignments were generated using the ClustalW option in MEGA 7. Phylogenetic trees were constructed with the maximum likelihood (ML) method

(Felsenstein 1981) based on a GTR+G+I model. Evolutionary divergence between sequences and percent similarity was estimated using the Maximum Composite Likelihood model. The aligned 16S rRNA, protein-coding housekeeping and symbiotic gene sequence fragments were manually trimmed to remove overhangs, and analyzed individually as well as in a concatenated fashion. The consistency of the tree topology was anticipated by conducting a bootstrap test with 1000 pseudo-replicates.

Nodulation tests and host range studies

Some sequenced rhizobial strains were assessed for their capacity to nodulate *C. pumila* under glass house conditions as described earlier (Gehlot *et al*., 2012, 2013; Sankhla *et al*., 2017). Rhizobial strains were re-isolated from nodules and compared with the parental strains. For host range studies, selected strains from *C. pumila* were cross-inoculated onto the crop legumes *Vigna radiata*, *Cyamopsis tetragonoloba*, *Glycine max* and also onto some wild legumes such as species of *Vachellia*, *Senegalia*, *Leucaena*, *Mimosa*, *Prosopis* and *Tephrosia*.

Environmental differentiation of the geographical regions of occurrence of rhizobial strains

The ENVIREM (Environmental Rasters for Ecological Modeling, http://envirem.github.io/) and the ISRIC (International Soil Reference and Information Centre, SoilGrids v0.5.8) datasets were used to analyse the influence of climate and soil on the prevalence of the *C. pumila* rhizobial strains. Generic grids of the ENVIREM dataset comprising of 13 variables with a spatial resolution of 30 arc seconds (~1 km) were downloaded from the archives of Deep Blue Data at the University of Michigan (https://deepblue.lib.umich.edu/data). The ISRIC dataset comprising of two variables viz., soil organic carbon stock and soil pH with a spatial resolution of ~1 km were downloaded from https://soilgrids.org. These datasets comprising of a set of 15 environmental variables (Table S4A) are relevant to species ecological and physiological processes (Title and Bemmels 2017).

Environmental differentiation of the geographical regions wherein the different *C. pumila* rhizobial strains in India were isolated was also performed using principal component analysis (PCA) based on elevation and 19 bioclimatic variables (Table S4B) (O'Donnell and Ignizio 2012). These bioclimatic variables are derivatives of monthly temperature and precipitation values, and comprise annual mean temperature and precipitation, mean diurnal temperature

range, annual temperature range, temperature and precipitation seasonality, isothermality, maximum temperature of warmest month, minimum temperature of coldest month, precipitation of driest and wettest month, and mean temperature and precipitation of wettest, driest, warmest and coldest quarter. They are relevant to the persistence of the selected strains/species. The temperature-related variables describe the thermal tolerance of the strain/species while the precipitation-related variables describe the water availability. These variables have a global coverage and are available at a spatial resolution of 30 arc seconds (~1 km), and can be downloaded from the worldclim database (www.worlclim.org). Digital elevation data of 90 m spatial resolution were downloaded from CGIAR-CSI (http://srtm.csi.cgiar.org, Jarvis *et al*., 2008). Those environmental variables pertaining to India were downloaded and resampled to a spatial resolution of one km using ArcGIS software.

Analysis of environmental data

Environmental data pertaining to the soil and climate of the sampling locations were extracted in '.txt' format from the ENVIREM and ISRIC datasets using ArcGIS software. For delineating spatial extents to extract environmental data polygons were constructed comprising a 10 km buffer area around each sampling location. These polygons delineated the spatial extent of the sampling region, and were later used to clip the raster data on elevation and bioclimatic variables using ArcGIS software. The clipped individual raster layers were combined using ArcGIS, and exported as a text file that contained the data of all the individual layers. This procedure was done for the climatic zones in each of the states.

Comparing environmental backgrounds of the sampled regions

A PCA was performed to compare the environmental backgrounds of the sampled regions using PAST software (ver. 3.17) (Hammer *et al*., 2001). To obtain an optimal solution 1000 bootstrap runs were madez

RESULTS

Nodulation status and nodule structure

Nodules of *C. pumila* excavated from various sampling sites (Fig. 1, S2 and Tables S1–S2) were all of the indeterminate type (Fig. S1F–G), and the multi-branched nodules were similar to those shown in Gehlot *et al*. (2012). There were fewer nodules per plant on the roots examined from soil in the alkaline and arid regions (*e.g.* RJ) in comparison to the acidic regions (*e.g.* ME).

Nodules were mainly present on secondary, tertiary and lateral roots in all the plants and were not seen on primary roots or at the collar region. The length of the nodules ranged from 2 to 20 mm. The general profile of the internal structure of the nodules from plants growing in both RJ (nodulated by *Ensifer*) (Fig. 2A) and ME (nodulated by *Bradyrhizobium*) (Fig. 2B) was similar in spite of the different rhizobial endophytes, and they resembled that recently reported for Brazilian *Chamaecrista* species nodulated by *Bradyrhizobium* (dos Santos *et al*., 2017). More detailed light microscopy (Fig. 2C) and TEM (Fig. 2D) showed that the numerous infected cells in the central zone were packed with bacteroids expressing the NifH protein of nitrogenase (Fig. 2D). Using the JIM5 probe it could be seen that the nodule cell walls contained pectin, as expected (Fig. 2E), but this was also the case with the symbiosome membranes surrounding the bacteroids, although the signal was much less than was observed on the thick cell walls of the invasive infection threads (Fig. 2F).

Origin of Chamaecrista pumila *rhizobial isolates*

More than 150 bacterial isolates were obtained and purified from the root nodules of *C. pumila* grown in soils with different pH (acidic-neutral-alkaline) values collected from six states (RJ, GJ, UT, TN, JH and ME) of India. From each site a minimum of 4–6 rhizobial isolates were obtained from nodules of 2–3 plants. More than 100 purified isolates were stored at 4^{o}C. The soil from ME (Shillong) was strongly acidic with pH 4.5–5.5; soils of TN (Pudukkottai), JH (Bokaro) and UT (Bhimtal) were near neutral (pH 6.6–7.0), whereas the soils of RJ (Western) and GJ (Valsad) were strongly to moderately alkaline (pH 8.0–9.0). All the rhizobial strains isolated from acidic soil were slow-growing taking about 4–6 days to grow at 28^{o}C, whereas the strains isolated from the alkaline soils of Western RJ were both slow- and fast- growing in equal proportions (Fig. S3). Other than in the Thar Desert (RJ) and Shillong (ME), the number of sampling sites and the area covered was limited in the present study. Therefore, from states other than RJ and ME, the types of rhizobia associated with *C. pumila* were only ascertained through trap experiments. In the limited sampling area from the alkaline soils of GJ only fast-growing strains were isolated. In the neutral soils of TN both slow- (60%) and fast- (40%) growing strains were found whereas in the limited sampling from JH and UT only slow-growing strains were isolated (Fig. S3).

Phylogenetic analysis of rhizobial strains nodulating Chamaecrista pumila *using housekeeping genes*

In the 16S rRNA gene phylogeny of seven (CP7, CP8, CP11, CP40, CP42, CPG48 and CPTN45) *Ensifer* strains isolated from different states (RJ, TN and GJ), four distinct genotypes were observed (Fig. S4), of which four representative strains (CP7, CP11, CPTN45 and CPG48) were selected for MLSA using four (*glnII*, *atpD*, *recA* and *dnaK*) protein-coding housekeeping genes. The four *Ensifer* strains isolated from different agro-climatic regions of India formed four distinct MLSA types in the four (*glnII-atpD-recA-dnaK*) (Fig. S5) and five genes (*rrs-glnII-atpD-recA-dnaK*) (Fig. 3) concatenated phylogenies. Three of the strains (CP7, CPG48 and CPTN45) formed distinct lineages within *Ensifer*, whereas strain CP11 clustered with *Tephrosia-Ensifer* strains previously isolated from the Thar Desert by Tak *et al.* (2016). The GenBank accession numbers for housekeeping and symbiotic genes of *C. pumila Bradyrhizobium* and *Ensifer* strains are given in Table S5.

Of the slow-growing *Bradyrhizobium* strains isolated from root nodules of *C. pumila* 26 were analysed for their 16S rRNA gene phylogeny (Fig. S6). It was observed that the 16S rRNA phylogeny had poor resolution, as the strains clustered with type strains, and the tree topology was not supported by strong bootstrap values. However, the phylogeny based on the housekeeping gene *recA* gave better resolution in terms of the phylogenetic positions of the 28 strains studied (*i.e.* the 26 strains used in the 16S rRNA gene phylogeny plus strains CP18 and CP28). Based on various novel clusters/groups and lineages formed in the *recA* phylogeny (Fig. S7) 19 *Bradyrhizobium* strains (Fig. S8) were selected for MLSA using three additional protein-coding housekeeping genes (*glnII*, *atpD* and *dnaK*). In the *rrs-glnII-recA-dnaK-atpD* concatenated phylogeny (Fig. 4) the 19 strains formed nine distinct MLSA types. The word "type" is used here as a general term for both the clusters and the lineages formed in the phylogenetic trees and does not infer that the strains within these lineages have identical nucleotide sequences. Eight MLSA types (I-VIII) constituting a total of 16 strains formed five novel clades/clusters, and three novel lineages within the *Bradyrhizobium* Mega clade-I, while the remaining three strains were in *Bradyrhizobium* Mega clade-II. MLSA types-I, II and III were close to *B. yuanmingense*; these included strains CP19, CP24 from RJ and strain CPTN33 from TN in a distinct clade (MLSA type-I), and the single strain lineages of CPUT49 (Bhimtal, UT) (MLSA type-II), and CPJH29 (Bokaro, JH) (MLSA type-III). The *C. pumila* strains isolated

from Shillong (ME) were genetically diverse consisting of six MLSA types (IV to IX). MLSA types-IV and V were related to each other and comprised strains CPS1 and CPS3 (MLSA type-IV), and strains CPS10, CPS30 and CPS48 (MLSA type-V). MLSA type-VI was a single lineage of strain CPS39 clustering close to *B. manausense*, MLSA type-VII was a novel clade containing strains CPS35, CPS38 and CPS42, and MLSA type-VIII was close to *B. japonicum* and consisted of strains CPS6 and CPS40. MLSA type-IX, constituting strains CPS12, CPS19 and CPS41 from ME, was the only genotype of *C. pumila* symbionts in Mega clade-II.

Phylogenetic analysis of Chamaecrista pumila *symbiont strains based on their symbiotic genes*

The symbiosis-essential genes (*nodA* and *nifH*) of four representative *Ensifer* strains (CP7, CP11, CPG48 and CPTN45) were amplified and sequenced. Incongruence was observed between core (housekeeping) and symbiotic (accessory) gene phylogenies of the *Ensifer* strains nodulating *C. pumila*. The four *Ensifer* strains formed three distinct symbiotic types in phylogenies based on *nodA* (Fig. S9), *nifH* (Fig. S10) and concatenated *nodA-nifH* (Fig. 5) gene sequences. *Ensifer* strains CP7 (isolated from Nagaur, RJ) and CPTN45 (isolated from Pudukkottai, TN) clustered together (*sym* type-I). The strain CPG48 (*sym* type-III) isolated from GJ formed a distinct symbiotic lineage close to *E. fredii*. All these *Ensifer* strains shared close similarity with *E. terangae* on the basis of core gene phylogenies and harbour novel *sym* genes closely related to *sym* genes of the broad host range strain *E. fredii* NGR234. The strain CP11 (*sym* type-II) was identical to the *E. aridi* (TW10) group of strains on the basis of core and accessory gene phylogenies. It should be noted that the *C. pumila* strains were quite distant from the Indian mimosoid-nodulating *Ensifer* strains, and clustered closer to those from papilionoid legumes, particularly *Tephrosia* species (Fig. 5).

In the symbiotic gene phylogenies, the 19 *Bradyrhizobium* strains from *C. pumila* clustered into 10 *sym* (*nodA* and *nifH*) types (Fig. 6 and Fig. 7) which were more or less similar to the MLSA types in the concatenated housekeeping gene phylogenies (Fig. 4 and Fig. S8). In the *nodA* phylogeny four *Bradyrhizobium* strains (CP19, CP24, CPTN33 and CPUT49) (*nodA* type-I) showed maximum similarity (97.9–99%) with *B. yuanmingense* and formed separate lineages close to it, as observed in the housekeeping gene phylogeny (Fig. 4). The *nodA* of strain CPJH29 (*nodA* type-II) was divergent from the above mentioned four *C. pumila Bradyrhizobium* strains, forming a novel lineage and sharing similarity (91.7%) to three strains (*B. daqingense*, *B. huanghuaihaiense* and *B. japonicum*) clustering together. Strain CPS39 (*nodA* type-III) was

divergent from all these strains and shared only 82.3% similarity with *B. yuanmingense* and strains in the *B. japonicum* clade. The other 10 (CPS1, CPS3, CPS6, CPS10, CPS30, CPS35, CPS38, CPS40, CPS42 and CPS48) *C. pumila Bradyrhizobium* strains from Shillong clustered into five different *nodA* types (*nodA* type-IV to VIII). Of which six strains (CPS1, CPS3, CPS35, CPS38, CPS40 and CPS42) shared maximum similarity (ranging from 95.2–100%) with *B. arachidis* whereas three strains of *nodA* type-VIII (CPS10, CPS30 and CPS48) formed a novel clade close to *B. forestalis* with 89.2% similarity. The single strain CPS6 shared similarity (95.7%) with both *B. arachidis* and *B. forestalis*. The remaining three (CPS12, CPS19 and CPS41) strains (*nodA* type-IX and X) from Shillong shared highest percentage similarities with *B. elkanii* (ranging from 94.1–99.5%), which is in accordance with their housekeeping gene phylogeny (Fig. 4).

The topology of the *nifH* phylogeny of the *Bradyrhizobium* strains from *C. pumila* was generally similar to the *nodA* phylogeny. Four strains (CP19, CP24, CPTN33 and CPUT49) (*nifH* type-I) showed maximum similarity (96.4–98.8%) with *B. yuanmingense*, but strain CPJH29 (*nifH* type-II) was divergent and shared only 92.7% similarity with *B. yuanmingense* and strains of the *B. japonicum* clade. The position and percentage similarity of strain CPS30 varied in the *nodA* and *nifH* phylogenies; although its *nodA* was closest to *B. forestalis* its *nifH* sequence showed more similarity (96.5%) with *B. yuanmingense*. Strain CPS39, forming the *nifH* type-III, showed close similarity (94%) with *B. centrosemae*. Strains of *nifH* type-IV (CPS1 and CPS3) shared maximum similarity (95.9%) with *B. subterraneum* and *B. vignae*. Strains of *nifH* type-V (CPS35, CPS38 and CPS42) and *nifH* type-VI (CPS40) were identical to *B. arachidis*, whereas strain CPS6 (*nifH* type-VII) shared similarity (98.3%) with both *B. arachidis* and *B. forestalis*, and strains CPS10 and CPS48 (*nifH* type-VIII) shared 94.6% similarity to *B. forestalis* and *B. pachyrhizi*. The same three strains in *Bradyrhizobium* Mega clade-II for the housekeeping genes and *nodA* phylogenies constituted *nifH* type-IX (CPS41) and *nifH* type-X (CPS12 and CPS19), and these shared 100% and 97.7% similarity, respectively, with *B. elkanii*.

Host authentication and cross-inoculation of rhizobial strains from Chamaecrista pumila

Plant nodulation tests were conducted to analyze the host range of the diverse rhizobial strains nodulating *C. pumila* (Table 1). All the *Bradyrhizobium* and *Ensifer* strains effectively nodulated their host and also the crop legume *V. radiata*, but the *Bradyrhizobium* strains isolated from *C.*

pumila in different agro-climatic conditions showed variation in their nodulation response on *G. max*. Five strains (CP19, CP24, CPJH29, CPTN33 and CPUT49) showing close similarity with *B. yuanmingense* could nodulate *G. max* whereas the other five *Bradyrhizobium* strains (CPS1, CPS6, CPS12, CPS19 and CPS30) isolated from acidic soils in Shillong (ME) failed to nodulate it. None of the tested *Bradyrhizobium* strains could nodulate another crop species, *C. tetragonoloba*, although it was nodulated by *Ensifer* (CP7 and CP11) strains isolated from *C. pumila*. Both wild species of *Tephrosia* tested for cross-nodulation, *T. villosa* and *T. wallichii* (Papilionoideae), were effectively nodulated by all the tested *Bradyrhizobium* and *Ensifer* strains. When the *Bradyrhizobium* strains were tested on various mimosoid legumes, they failed to nodulate *Leucaena leucocephala*, *Mimosa hamata*, *Senegalia senegal* and species of *Vachellia* (*V. jacquemontii*, *V. leucophloea* and *V. tortilis*). However, native *Prosopis cineraria* plants were nodulated by both the *Bradyrhizobium* and *Ensifer* strains in the present study. Among the various strains tested, only *Bradyrhizobium* strains (CP19 and CP24) and *Ensifer* sp. CP11 effectively cross-nodulated the invasive *Prosopis juliflora*. *Ensifer* sp. CP11 also nodulated other mimosoid legumes, such as (invasive) *L. leucocephala*, (introduced) *V. tortilis* and (native) *V. leucophloea*, but it failed to nodulate the closely related Indian endemic *V. jacquemontii*. Although the tested *Ensifer* strains were generally more capable than the *Bradyrhizobium* strains of nodulating mimosoid legumes, they still failed to nodulate either (native) *M. hamata* or *S. senegal*.

Principal Component Analysis

In PCA done using the environmental variables extracted for the location of the *C. pumila* rhizobial strains from the ENVIREM and ISRIC datasets the first two components explained ~91% of the total variation in the dataset, where the first component accounted for 80.8% of the total variation while the second component accounted for 10.3% of the total variation (Fig. 8, Table S6A). The environmental variables having high correlations with the first PC axis are soil pH, soil organic carbon, PETWQ, PETWtQ, PETCQ, MTWQ, CMI, AI, APET, while the variables having relatively high correlations with the second PC axis are thermicity index and continentality (Table S6A). The first principal component axis clearly segregates ME from RJ, while the second axis segregates TN and GJ from UT. High soil organic carbon (SOC) and climatic moisture index (CMI) were the environmental variables most associated with the

partitioning of the ME strains separately from the other populations. While strains in RJ have positive associations with high soil pH, annual potential evapotranspiration rate, and aridity index (Fig. 8). The *C. pumila* rhizobial strains from TN and GJ are positively associated with thermicity index, while those from UT are positively associated with continentality.

In PCA based on elevation and bioclimatic variables the first two components explained ~82% of the total variation in the dataset, where the first component (PC1-elevation) accounted for 58.1% of the total variation while the second component (PC2-annual mean temperature) accounted for 24.4% of the total variation (Fig. S11, Table S6B). The first principal component axis clearly segregates the states of ME and UT from RJ, while the second axis segregates TN and GJ from UT and RJ (Fig. S11). The locations in ME and UT were characterized by elevation and precipitation related variables, while the locations in RJ were characterized by temperature-related variables. The locations in the states of TN and GJ were characterized by isothermality.

DISCUSSION

Nodulation and nodule structure of Chamaecrista pumila

Chamaecrista pumila was first reported to nodulate by Gehlot *et al*. (2012), but only nodule morphology was described and no rhizobial symbionts were characterized. The structure of nodules on *C. pumila*, regardless of the symbiont type (*Bradyrhizobium* or *Ensifer*), were generally similar to those described on herbaceous *Chamaecrista* species from other parts of the world. It was previously noted that the bacteroids in nodules on perennial woody species of *Chamaecrista* (large shrubs and small trees) were contained within fixation threads (FTs), which are a common (even defining) feature of nodules on all caesalpinioid (s.s.) species outside *Chamaecrista* (de Faria *et al*., 1987; Naisbitt *et al*., 1992; Fonseca *et al*., 2012). These FTs contain pectin which is a component of the plant cell wall, as evidenced by immunogold labelling with JIM5 (Fonseca *et al*., 2012). In the case of *C. pumila*, the bacteroids were not enclosed in a thick cell wall, but in a symbiosome membrane similar to that seen in nodules on other non-woody, herbaceous *Chamaecrista* species (Naisbitt *et al*., 1992; dos Santos *et al*., 2017). However, the symbiosome membrane was also labelled with JIM5, which shows that bacteroids even in the herbaceous *Chamaecrista* species are surrounded to some degree by pectin-containing material, suggesting that the presence or not of FTs is not such an absolute feature, but may be a matter of the degree to which the symbiosome membrane is impregnated with pectin (and possibly other cell wall components) *i.e.* the larger and woodier the species the

more pectin is contained within the membrane, such that its appearance means that it can be effectively termed an FT.

Chamaecrista pumila *is nodulated by diverse* Bradyrhizobium *and* Ensifer *strains*

Most of the strains isolated from *C. pumila* nodules in the present study were *Bradyrhizobium*, but this is the first report of nodulation of *C. pumila* by *Ensifer*, and to our knowledge is the first report of any *Chamaecrista* species being nodulated by this rhizobial type. Almost all previous reports on symbionts of native *Chamaecrista* species in both the Old (Beukes *et al*., 2016) and the New World (dos Santos *et al*., 2017) have indicated that they were almost exclusively nodulated by species of *Bradyrhizobium*. It is known that edaphic and climatic factors play a major role in the distribution of legume symbionts in native ecosystems (Lemaire *et al*., 2015; Pires *et al*., 2018). For example, the PCA shows that soil pH is clearly a factor in the occurrence of diverse strains of *Ensifer* nodulating *C. pumila* in the alkaline soil of RJ; this rhizobial type has been frequently isolated from several native legumes in this region (Gehlot *et al*., 2012, 2013, 2016; Tak *et al*., 2013, 2016; Le Queré *et al*., 2017; Sankhla *et al*., 2017, 2018; Choudhary *et al*., 2017, 2018). However, *Ensifer* strains were not the only symbionts of *C. pumila* in the alkaline soils of RJ; *Bradyrhizobium* MLSA type-I strains that were related to *B. yuanmingense* were also isolated. In addition, *C. pumila* growing in the neutral pH soils of UT (CPUT49), JH (CPJH29), and TN (CPTN33) yielded similar *B. yuanmingense*-like symbionts. This is not too surprising, as *B. yuanmingense*, has been isolated frequently from soybean and *Vigna* species growing in alkaline to weakly acidic soils in other parts of India (Appunu *et al*., 2008, 2009a, 2009b; Vinuesa *et al*., 2008).

Bradyrhizobia were the only symbiont types isolated from *C. pumila* growing in Shillong (ME), but in contrast to the alkaline-neutral soils of RJ, TN, UT and JH, no strains were related to *B. yuanmingense*. Instead, these strains which were isolated from high altitude, high rainfall and acidic soils of the Eastern Himalayan region were phylogenetically diverse forming novel clades and lineages across both *Bradyrhizobium* Mega-clades. However, there was a clear dominance of the *Bradyrhizobium* Mega clade-I type of strains as microsymbionts of *C. pumila*, with only three (CPS12, CPS19 and CPS41) forming a novel sub-clade within Mega clade-II. The Shillong-*Bradyrhizobium* strains were genetically diverse from established type strains, as well as from *Bradyrhizobium* strains associated with native *Chamaecrista* species growing in

acidic soils of South Africa (Beukes *et al*., 2016) and Brazil (dos Santos *et al.*, 2017). The Shillong *C. pumila Bradyrhizobium* strains were also divergent from *Bradyrhizobium* strains recently isolated from *Eriosema chinense* and *Flemingia vestita* (tribe Phaseoleae) growing in similar acidic soils in the same region of ME (Ojha *et al*., 2017). Taken together, these data suggest that the widespread species *C. pumila* in India is not restricted to the "traditional" symbionts of *Chamaecrista* which have been reported in other parts of the tropics and sub-tropics *e.g.* none of them were closely related to the bradyrhizobia which nodulate the genus in its main centre of radiation in Brazil (dos Santos *et al*., 2017).

The selection of compatible rhizobia by the host legume depends upon the molecular dialogue between the two symbiotic partners, but it is also influenced by the local ecological factors such as soil pH (Yang *et al*., 2001; this study), soil nutrient availability (Pires *et al*., 2018) and precipitation. Pires *et al*. (2018) in their study on *Mimosa* symbionts in central Brazil noticed a very strong effect of soil pH on symbiont preference to the extent that endemic *Mimosa* species, such as *M. claussenii*, normally found to be nodulated (with *Paraburkholderia*) in low pH soils could not nodulate (with *Rhizobium*) in soils with neutral-alkaline pH, whereas more widespread *Mimosa* species were capable of nodulating with a greater variety of symbionts in soils with a wide range of pH levels. In terms of the composition of the microsymbiont community of *C. pumila* in India, it resembles widespread *Mimosa* species in Brazil, in that it is niche-specific and changes according to soil pH, resulting in the total dominance of *Bradyrhizobium* strains in acidic soils, but with *Ensifer* strains forming a substantial minority of symbionts in neutral-alkaline soils.

Symbiotic diversity and host range of Chamaecrista pumila *rhizobial strains*

The phylogeny of the *sym* (*nodA* and *nifH*) genes of the *Ensifer* strains isolated from *C. pumila* growing in the alkaline and neutral soils of three Indian states (RJ, TN and GJ) were quite incongruent with their housekeeping gene phylogeny, and were close to, but still substantially divergent from, the *sym* genes of *E. fredii* and other *Ensifer* species which nodulate soybean. The phylogenetic discordance between the core and *sym* genes in these *Ensifer* strains suggests plasmid-mediated lateral transfer of the *sym* genes (Haukka *et al*., 1998). All four of the strains analyzed were nested within a clade containing strains isolated from *Tephrosia* species in RJ (Tak *et al*., 2016), indicating the wide distribution of these novel *sym* genes within this region of

India. It is particularly interesting in terms of co-evolution of the symbiotic partners that the *sym* genes of the *Ensifer* strains isolated from *C. pumila* (Caesalpinioideae) clustered with the *sym* genes specific to local Papilionoideae (*Tephrosia*) (Tak *et al*., 2016) rather than to the local mimosoids (*Mimosa*, *Senegalia* and *Vachellia*) (Gehlot *et al*., 2013; Sankhla *et al*., 2017; Choudhary *et al*., 2017, 2018). In accordance with their *sym* gene phylogenies, the *C. pumila Ensifer* strains nodulated *Tephrosia* species, as well as several other promiscuous crop (*Vigna*) and invasive wild legumes (*L. leucocephala*, *P. juliflora* and *V. tortilis*). Out of this group of four *Ensifer* strains, of particular note is CP11, which is identical to *E. aridi* (Tak *et al*., 2016; Le Queré *et al*., 2017) in both its core and *sym* gene phylogenies, and like other *E. aridi* strains it is promiscuous, effectively nodulating several papilionoids (both crop and wild) and mimosoids.

The *B. yuanmingense*-like strains (CP19, CP24, CPTN33 and CPUT49) clustered together in their *sym* and housekeeping gene phylogenies. They were quite promiscuous, capable of nodulating several papilionoid hosts, including soybean, *Vigna* and *Tephrosia* species, but they failed to nodulate the mimosoid species tested (except for the *Prosopis* species). The present study, taken together with earlier ones, illustrates the widespread occurrence in India of *B. yuanmingense*-like strains, especially in neutral to alkaline soils. However, it should be stressed that such strains are not confined to neutral to alkaline soils and are also found in mildly acidic ones (Appunu *et al*., 2009a), although never in the more acidic soils of ME (Ojha *et al*., 2017; this study).

The genetically diverse *Bradyrhizobium* strains isolated from *C. pumila* growing in acidic soils in Shillong (ME) harbored diverse *sym* genotypes. Most of the Shillong strains (CPS1, CPS3, CPS6, CPS10, CPS30, CPS35, CPS38, CPS40, CPS42 and CPS48) were highly divergent from *B. arachidis* (isolated from *A. hypogaea* in China; Wang *et al*., 2013) in their housekeeping gene phylogeny, but harbored *sym* genes that were similar to or identical to it or to *B. forestalis* (isolated from *Inga* and *Swartzia* species in Brazil; Martins da Costa *et al*., 2018). Indeed, horizontal gene transfers are known to have generated *Bradyrhizobium* strains which carry mosaic combinations of symbiotic and core genes (Parker and Rousteau 2014). The remaining Shillong strains, CPS12, CPS19 and CPS41, clustered within Mega clade-II both in their core and *sym* gene phylogenies, with their symbiotic genes being most closely related to *B. elkanii* from soybean (Kuykendall *et al*., 1992). These were the only *C. pumila* strains from the present study in this Mega clade-II, and they had *sym* genes similar to other Shillong bradyrhizobia, such

as strains EHJO8 and EHNEHU6 from the papilionoid legume *Eriosema* (Ojha *et al*., 2017). As with the *B. yuanmingense*-like strains isolated from *C. pumila* in other parts of India, the ME bradyrhizobia could nodulate *Vigna* and *Tephrosia* species and failed to nodulate the mimosoid species tested (except for *Prosopis cineraria*), but unlike the *B. yuanmingense*-like strains none of them nodulated soybean, which contrasts with the aforementioned *Bradyrhizobium* strains from *E. chinense* and *F. vestita* (Ojha *et al*., 2017), and is particularly surprising in the case of the Mega clade-II strains, CPS12, CPS19 and CPS41, as they had *nodA* genes that were similar to *B. elkanii*.

The *nodA* and *nifH* gene sequences of the *C. pumila Bradyrhizobium* strains were mostly congruent supporting their monophyletic origin, except for a few strains showing differences. There was no correlation between the phylogenetic affinities of the *C. pumila*-nodulating *Bradyrhizobium* strains based on individual and concatenated core gene sequences with that of their symbiotic genes, except for a few *B. yuanmingense* strains that maintained consistency for most of the studied gene sequences. These results suggest contrasting evolutionary histories of core and symbiotic genes for most of the *Bradyrhizobium* strains from *C. pumila*. The congruence of *nodA* and *nifH* trees and the high symbiotic diversity reported in this and previous studies (Moulin *et al*., 2004; Aserse *et al*., 2012) suggests that these genes in *Bradyrhizobium* strains are spread both by vertical and horizontal gene transfer, and then maintained. Taken together with the cross-inoculation results, the data from the present study demonstrate that *C. pumila* has no particular specificity for a particular microsymbiont type, and is capable of adopting the symbionts of other native legumes growing in a particular niche.

The acidic soil-specific *sym* genes of *Bradyrhizobium* strains associated with *Chamaecrista* species in India, Africa and Brazil are phylogenetically diverse. These results indicate selection of distinct microsymbionts harboring specific *sym* genes by host legumes depends upon soil pH as well as on various other geographic-climatic factors. *Chamaecrista pumila* has a long history of occurrence throughout India, and its associated rhizobial microsymbionts have co-evolved with it in different niches. The high altitude, high precipitation rate and the acidic soils of Shillong favored the evolution of diverse genotypes of *Bradyrhizobium* strains. However, it should be noted that the *C. pumila* symbionts in India are still quite closely related to rhizobia which nodulate other legumes in the regions in which it lives *i.e.* it has recruited the local rhizobial micro-flora for its symbiotic requirements rather than adopt

"*Chamaecrista*-specific" *Bradyrhizobium* symbionts like their cousins in Brazil, which is the centre of radiation of the genus (dos Santos *et al*., 2017). In this respect *C. pumila* resembles *Mimosa himalayana* and *M. hamata*, which have adopted local Indian mimosoid-nodulating *Ensifer* strains as symbionts rather than being associated with the (Beta) rhizobial strains of their close relatives in South America (Gehlot *et al*., 2013). The core and symbiotic gene sequences of *Chamaecrista* sp. microsymbionts reported from three different continents (Africa, America and Asia {this study}) are significantly different from each other, supporting the likelihood of divergent evolution due to geographical separation followed by adaptation to local edaphic and climatic conditions, which also influences the availability of compatible legume hosts. Although microsymbionts are peripatetic and cosmopolitan their global distribution is determined primarily by their adaptability to a particular habitat rather than by geographic contiguity.

ACKNOWLEDGEMENTS

Authors are thankful to the following funding agencies for financial assistance: University Grants Commission, New Delhi (UGC-JRF and UGC-SRF to S.R., UGC-SAPII-CAS-I, UGC-MRP F.No.43-129/2014 [SR]); Department of Science and Technology (DST)-SERB (SB/YS/LS-369/2013) and Department of Biotechnology (DBT), New Delhi for awarding Visiting Research Professor-North-Eastern Region (BT/49/NE/2014) to HSG.

Conflict of Interest: All the authors declare that there is no conflict of interest regarding the publication of this article.

REFERENCES

Altschul SF, Gish W, Miller W, *et al*. Basic local alignment search tool. J Mol Biol 1990;**215**:403–10.

Andrews M, Andrews ME. Specificity in legume-rhizobia symbioses. Int J Mol Sci 2017;**18**: 705.

Appunu C, N'Zoue A, Laguerre G. Genetic diversity of native bradyrhizobia isolated from soybeans (*Glycine max* L.) in different agricultural-ecological-climatic regions of India. Appl Environ Microbiol 2008;**74**:5991–6.

Appunu C, N'Zoue A, Moulin L, *et al*. *Vigna mungo*, *V. radiata* and *V. unguiculata* plants sampled in different agronomical-ecological-climatic regions of India are nodulated by *Bradyrhizobium yuanmingense*. Syst Appl Microbiol 2009a;**32**:460–70.

Appunu C, Sasirekha N, Prabavathi VR, *et al*. A significant proportion of indigenous rhizobia from India associated with soybean (*Glycine max* L.) distinctly belong to *Bradyrhizobium* and *Ensifer* genera. Biol Fertil Soils 2009b;**46**:57–63.

Aserse AA, Räsänen LA, Aseffa F, *et al*. Phylogenetically diverse groups of *Bradyrhizobium* isolated from nodules of *Crotalaria* spp., *Indigofera* spp., *Erythrina brucei* and *Glycine max* growing in Ethiopia. Mol Phylogenet Evol 2012;**65**:595–609.
Barneby RC. A new species of *Chamaecrista* Moench (Caesalpiniaceae) from interior Bahia, Brazil. Brittonia 1994;**46**:69–71.
Beukes CW, Stepkowski T, Venter SN, *et al*. Crotalarieae and Genisteae of the South African great escarpment are nodulated by novel *Bradyrhizobium* species with unique and diverse symbiotic loci. Mol Phylogenet Evol 2016;**100**:206–18.
Chaintreuil C, Boivin C, Dreyfus B, *et al*. Characterization of the common nodulation genes of the photosynthetic *Bradyrhizobium* sp. ORS285 reveals the presence of a new insertion sequence upstream of *nodA*. FEMS Microbiol Lett 2001;**194**:83–86.
Cheng HR, Jiang N. Extremely rapid extraction of DNA from bacteria and yeasts. Biotechnol Lett 2006;**28**:55–9.
Choudhary S, Meghwal RR, Sankhla IS, *et al*. Molecular characterization and phylogeny of novel diverse nitrogen fixing microsymbionts associated with *Vachellia* (*Acacia*) *leucophloea* in arid and semi-arid regions of Rajasthan. Indian Forester 2017;**143**:266–78.
Choudhary S, Tak N, Gehlot HS. Phylogeny and genetic diversity assessment of *Ensifer* strains nodulating *Senegalia* (*Acacia*) *senegal* (L.) Britton. in arid regions of Western Rajasthan, India. Microbiology 2018;**87**:127–42.
de Faria SM, McInroy SG, Sprent JI. The occurrence of infected cells, with persistent infection threads, in legume root nodules. Can J Bot 1987;**65**:553–8.
dos Reis Junior FB, Simon MF, Gross E, *et al*. Nodulation and nitrogen fixation by *Mimosa* spp. in the Cerrado and Caatinga biomes of Brazil. New Phytol 2010;**186**:934–46.
dos Santos JMF, Alves PAC, Silva VC, *et al*. Diverse genotypes of *Bradyrhizobium* nodulate herbaceous *Chamaecrista* (Moench) (Fabaceae, Caesalpinioideae) species in Brazil. Syst Appl Microbiol 2017;**40**:69–79.
Doyle JJ. Phylogenetic perspectives on the origins of nodulation. Mol Plant Microbe Interact 2011;**11**:1289–95.
Elliott GN, Chen WM, Chou JH, *et al*. *Burkholderia phymatum* is a highly effective nitrogen-fixing symbiont of *Mimosa* spp. and fixes nitrogen *ex planta*. New Phytol 2007;**173**:168–80.
Felsenstein J. Evolutionary trees from DNA-sequences-a maximum likelihood approach. J Mol Evol 1981;**17**:368–76.
Fonseca MB, Peix A, de Faria SM, *et al*. Nodulation in *Dimorphandra wilsonii* Rizz. (Caesalpinioideae), a threatened species native to the Brazilian Cerrado. PLoS One 2012;**7**: e49520.
Gaunt MW, Turner SL, Rigottier-Gois L, *et al*. Phylogenies of *atpD* and *recA* support the small subunit rRNA-based classification of rhizobia. Int J Syst Evol Microbiol 2001;**51**:2037–48.
Gehlot HS, Ardley J, Tak N, *et al*. High-quality permanent draft genome sequence of *Ensifer* sp. PC2, isolated from a nitrogen-fixing root nodule of the legume tree (Khejri) native to the Thar Desert of India. Stand Genomic Sci 2016;**11**:43.
Gehlot HS, Panwar D, Tak N, *et al*. Nodulation of legumes from the Thar desert of India and molecular characterization of their rhizobia. Plant Soil 2012;**357**:227–43.
Gehlot HS, Tak N, Kaushik M, *et al*. An invasive *Mimosa* in India does not adopt the symbionts of its native relatives. Ann Bot 2013;**112**:179–96.
Gyaneshwar P, Hirsch AM, Moulin L, *et al*. Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. Mol Plant Microbe Interact 2011;**24**:1276–88.

Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological Statistics Software Package for Education and Data Analysis. Palaeontologia Electronica 2001;**4**:9.
Haukka K, Lindstrom K, Young JPW. Three phylogenetic groups of *nodA* and *nifH* genes in *Sinorhizobium* and *Mesorhizobium* isolates from leguminous trees growing in Africa and Latin America. Appl Environ Microbiol 1998;**64**:419–26.
Irwin HS, Barneby RC. The American Cassinae: A synoptical revision of Leguminosae tribe Cassieae subtribe Cassinae in the New World. Mem N Y Bot Gard 1982;**35**:1–918.
Jarvis A, Reuter HI, Nelson EG. Hole-filled Seamless SRTM Data V4. International Centre for Tropical Agriculture (CIAT), from http://srtm.csi.cgiar.org. 2008.
Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 2016;**33**:1870–4.
Kuykendall LD, Saxena B, Devine TE, *et al*. Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov. Can J Microbiol 1992;**38**:501–5.
Lafay B, Burdon JJ. Molecular diversity of legume root-nodule bacteria in Kakadu National Park, Northern Territory, Australia. PLoS One 2007;**2**:e277.
Laguerre G, Nour SM, Macheret V, *et al*. Classification of rhizobia based on *nodC* and *nifH* gene analysis reveals a close phylogenetic relationship among *Phaseolus vulgaris* symbionts. Microbiology 2001;**147**:981–93.
Le Queré A, Tak N, Gehlot HS, *et al*. Genomic characterization of *Ensifer aridi*, a proposed new species of nitrogen-fixing rhizobium recovered from Asian, African and American deserts. BMC Genomics 2017;**18**:85.
Lemaire B, Dlodlo O, Chimphango S, *et al*. Symbiotic diversity, specificity and distribution of rhizobia in native legumes of the Core Cape Subregion (South Africa). FEMS Microbiol Ecol 2015;**91**:1–17.
Lewis GP, Schrire B, Mackinder B, *et al*. Legumes of the World. Kew, UK: Royal Botanic Gardens, Kew, 2005.
Lewis GP. Legumes of Bahia. London, UK: Royal Botanic Gardens, Kew, 1987.
Lewis GP. Tribe Cassieae. In: Lewis GP, Schrire B, Mackinder B, Lock M (Eds.) Legumes of the World. Royal Botanic Gardens, Kew, 2005;111–61.
LPWG, Legume Phylogeny Working Group. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon 2017;**66**:44–77.
Martens M, Dawyndt P, Coopman R, *et al*. Advantages of multilocus sequence analysis for taxonomic studies: a case study using 10 housekeeping genes in the genus *Ensifer* (including former *Sinorhizobium*). Int J Syst Evol Microbiol 2008;**58**:200–14.
Martins da Costa E, Azarias Guimarães A, Soares de Carvalho T, *et al*. *Bradyrhizobium forestalis* sp. nov., an efficient nitrogen-fixing bacterium isolated from nodules of forest legume species in the Amazon. Arch Microbiol 2018;**200**:743–52.
Menna P, Barcellos FG, Hungria M. Phylogeny and taxonomy of a diverse collection of *Bradyrhizobium* strains based on multilocus sequence analysis of the 16S rRNA gene, ITS region and *glnII, recA, atpD* and *dnaK* genes. Int J Syst Evol Microbiol 2009;**59**:2934–50.
Moreira FMS, Haukka K, Young JPW. Biodiversity of rhizobia isolated from a wide range of forest legumes in Brazil. Mol Ecol 1998;**7**:889–95.
Moulin L, Béna G, Boivin-Masson C, *et al*. Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. Mol Phylogenet Evol 2004;**30**:720–32.

Naisbitt T, James EK, Sprent JI. The evolutionary significance of the genus *Chamaecrista*, as determined by nodule structure. New Phytol 1992;**122**:487–92.
O'Donnell MS, Ignizio DA. Bioclimatic predictors for supporting ecological applications in the conterminous United States: U.S. Geological Survey Data Series 2012;**691**:10.
Ojha A, Tak N, Rathi S, *et al*. Molecular characterization of novel *Bradyrhizobium* strains nodulating *Eriosema chinense* and *Flemingia vestita*, important unexplored native legumes of the sub-Himalayan region (Meghalaya) of India. Syst Appl Microbiol 2017;**40**: 334–44.
Parker MA, Kennedy DA. Diversity and relationships of bradyrhizobia from legumes native to eastern North America. Can J Microbiol 2006;**52**:1148–57.
Parker MA, Rousteau A. Mosaic origins of *Bradyrhizobium* legume symbionts on the Caribbean island of Guadeloupe. Mol Phylogenet Evol 2014;**77**:110–5.
Parker MA. The spread of *Bradyrhizobium* lineages across host legume clades: from Abarema to Zygia. Microb Ecol 2015;**69**:630–40.
Parker MA. Divergent *Bradyrhizobium* symbionts on *Tachigali versicolor* from Barro Colorado Island, Panama. Syst Appl Microbiol 2000;**23**:585–90.
Parker MA. Legumes select symbiosis island sequence variants in *Bradyrhizobium*. Mol Ecol 2012;**21**:1769–78.
Peix A, Ramírez-Bahena MH, Velázquez E, *et al*. Bacterial associations with legumes. Crit Rev Plant Sci 2015;**34**:17–42.
Pires RC, dos Reis Junior FB, Zilli JE, *et al*. Soil characteristics determine the rhizobia in association with different species of *Mimosa* in central Brazil. Plant Soil 2018;**423**:411–28.
Sankhla IS, Meghwal RR, Choudhary S, *et al*. Molecular characterization of microsymbionts associated with root nodules of *Crotalaria burhia*, a native key stone legume species from the Thar Desert of India. Indian J Exp Biol 2018;**56**:373–84.
Sankhla IS, Tak N, Meghwal RR, *et al*. Molecular characterization of nitrogen fixing microsymbionts from root nodules of *Vachellia* (*Acacia*) *jacquemontii*, a native legume from the Thar Desert of India. Plant Soil 2017;**410**:21–40.
Somasegaran P, Hoben HJ. Methods in legume rhizobium technology. Springer, New York, 1994.
Sprent JI, Ardley JK, James EK. Biogeography of nodulated legumes and their nitrogen fixing symbionts. New Phytol 2017;**215**:40–56.
Sprent JI, Ardley JK, James EK. From North to South: a latitudinal look at legume nodulation processes. S Afr J Bot 2013;**89**:31–41.
Sprent JI. Bacteria Nodulating Legumes, in Legume Nodulation: A Global Perspective. Oxford, UK: Wiley-Blackwell, 2009.
Stepkowski T, Czaplinska M, Miedzinska K, *et al*. The variable part of the *dnaK* gene as an alternative marker for phylogenetic studies of rhizobia and related alpha Proteobacteria. Syst Appl Microbiol 2003;**26**:483–94.
Stepkowski T, Moulin L, Krzyzanska A, *et al*. European origin of *Bradyrhizobium* populations infecting lupins and serradella in soils of Western Australia and South Africa. Appl Environ Microbiol 2005;**71**:7041–52.
Tak N, Awasthi E, Bissa G, *et al*. Multi locus sequence analysis and symbiotic characterization of novel *Ensifer* strains nodulating *Tephrosia* spp. in the Indian Thar Desert. Syst Appl Microbiol 2016;**39**:534–45.

Tak N, Gehlot HS, Kaushik M, *et al*. Genome sequence of *Ensifer* sp. TW10; a *Tephrosia wallichii* (Biyani) microsymbiont native to the Indian Thar Desert. Stand Genomic Sci 2013;**9**:304–14.
Title PO, Bemmels JB. ENVIREM: An expanded set of bioclimatic and topographic variables increases flexibility and improves performance of ecological niche modeling. Ecography 2017;**41**:291–307.
Turner SL, Young JP. The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. Mol Biol Evol 2000;**17**:309–19.
Vinuesa P, Rojas-Jiménez K, Contreras-Moreira B, *et al*. Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent. Appl Environ Microbiol 2008;**74**:6987–96.
Wang R, Chang YL, Zheng WT, *et al*. *Bradyrhizobium arachidis* sp. nov., isolated from effective nodules of *Arachis hypogaea* grown in China. Syst Appl Microbiol 2013;**36**:101–5.
Weisburg WG, Barns SM, Pelletier DA, *et al*. 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 1991;**173**:697–703.
Yang J, Zhou Q, Zhou J. Effect of pH on nodulation of soybean rhizobia from Weifang and Huayuankou soils. Chin J Appl Ecol 2001;**12**:639–40.
Yao Y, Sui XH, Zhang XX, *et al*. *Bradyrhizobium erythrophlei* sp. nov. and *Bradyrhizobium ferriligni* sp. nov., isolated from effective nodules of *Erythrophleum fordii*. Int J Syst Evol Microbiol 2015;**65**:1831–37.
Yao Y, Wang R, Lu JK, *et al*. Genetic diversity and evolution of *Bradyrhizobium* populations nodulating *Erythrophleum fordii*, an evergreen tree indigenous to the southern subtropical region of China. Appl Environ Microbiol 2014;**80**:6184–94.

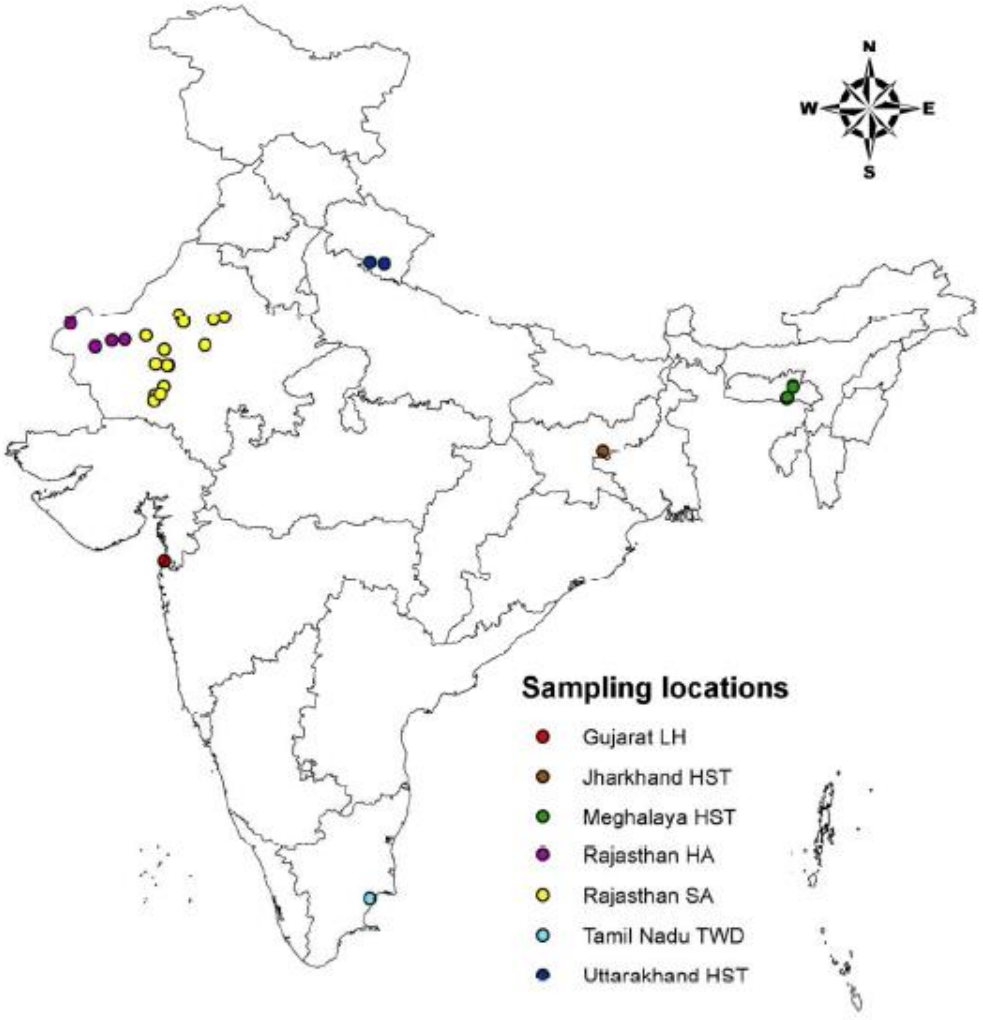

Fig. 1 Sampling locations of rhizobial strains in different states of India. The abbreviations against the state names can be expanded as tropical wet and dry (TWD), humid subtropical (HST), Low humid (LH), semi-arid (SA), and hyper-arid (HA).

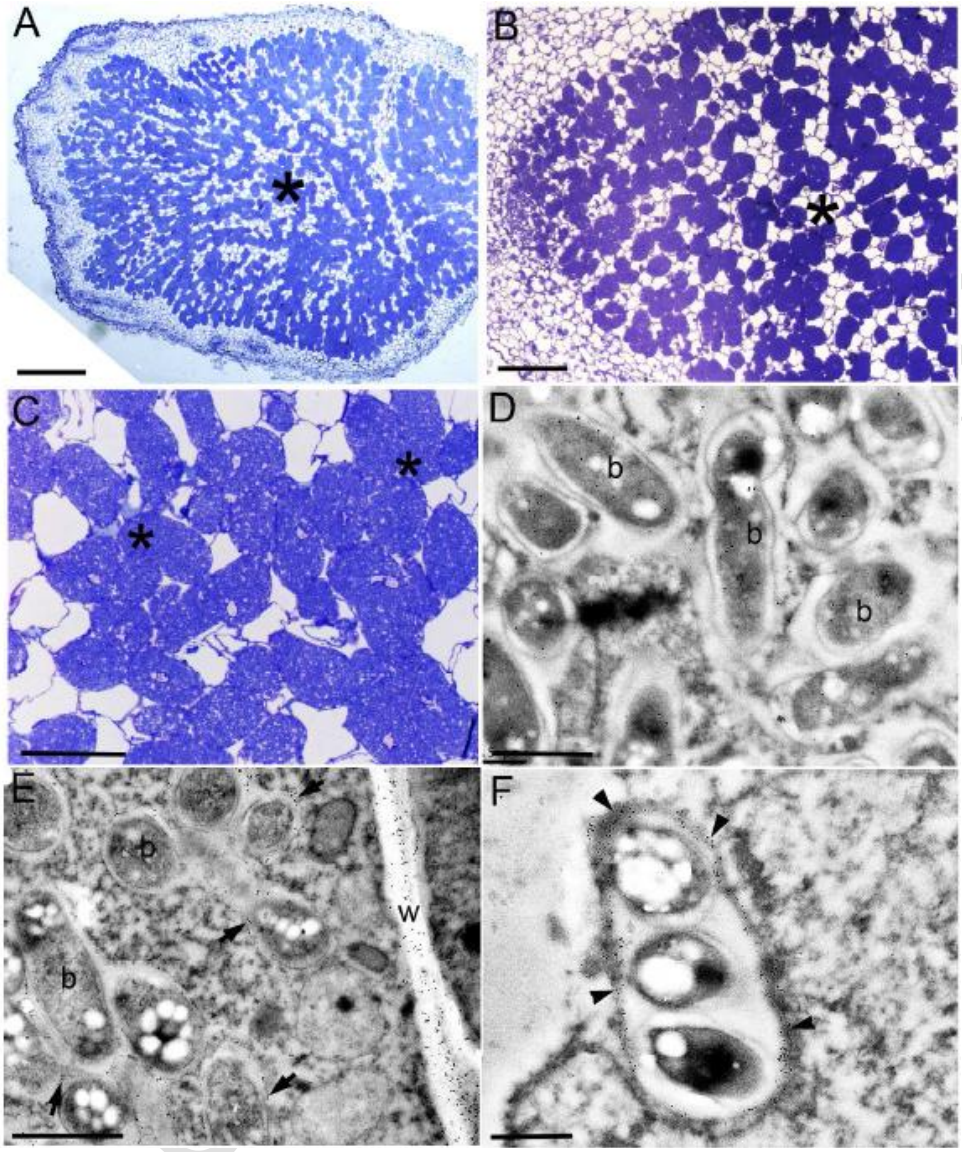

Fig. 2 Structure of *Chamaecrista pumila* nodules examined using light (A–C) and transmission electron microscopy (D–F).

(A) Profile of section of nodule from the Thar Desert (RJ). The infected zone is marked with *. Bar = 200 µm.

(B) Profile of section of nodule from Shillong (ME). The infected zone is marked with *. Bar = 50 µm.

(C) Higher magnification of infected cells packed with bacteria (*) in a nodule from the Thar Desert (RJ). Note the smaller, uninfected cells between the infected cells. Bar = 20 µm.

(D) Bacteroids (b) in an infected cell in a nodule from the Thar Desert (RJ). This section was immunogold labelled with an antibody against the NifH protein of nitrogenase; the labelling is visible as 15 nm diameter black dots on the bacteroids. Bar = 1 µm.

(E) Bacteroids (b) in an infected cell in a nodule from Shillong (ME). This section was immunogold labelled with JIM5, a monoclonal antibody recognizing pectin; the labelling is visible as abundant 10 nm diameter black dots on the plant cell wall (w), but there is also some sparser labelling surrounding the bacteroids (arrows). Bar = 1 µm.

(F) JIM5 labelling (arrowheads) of the cell wall of an infection thread in a nodule from the Thar Desert (RJ). Bar = 500 nm.

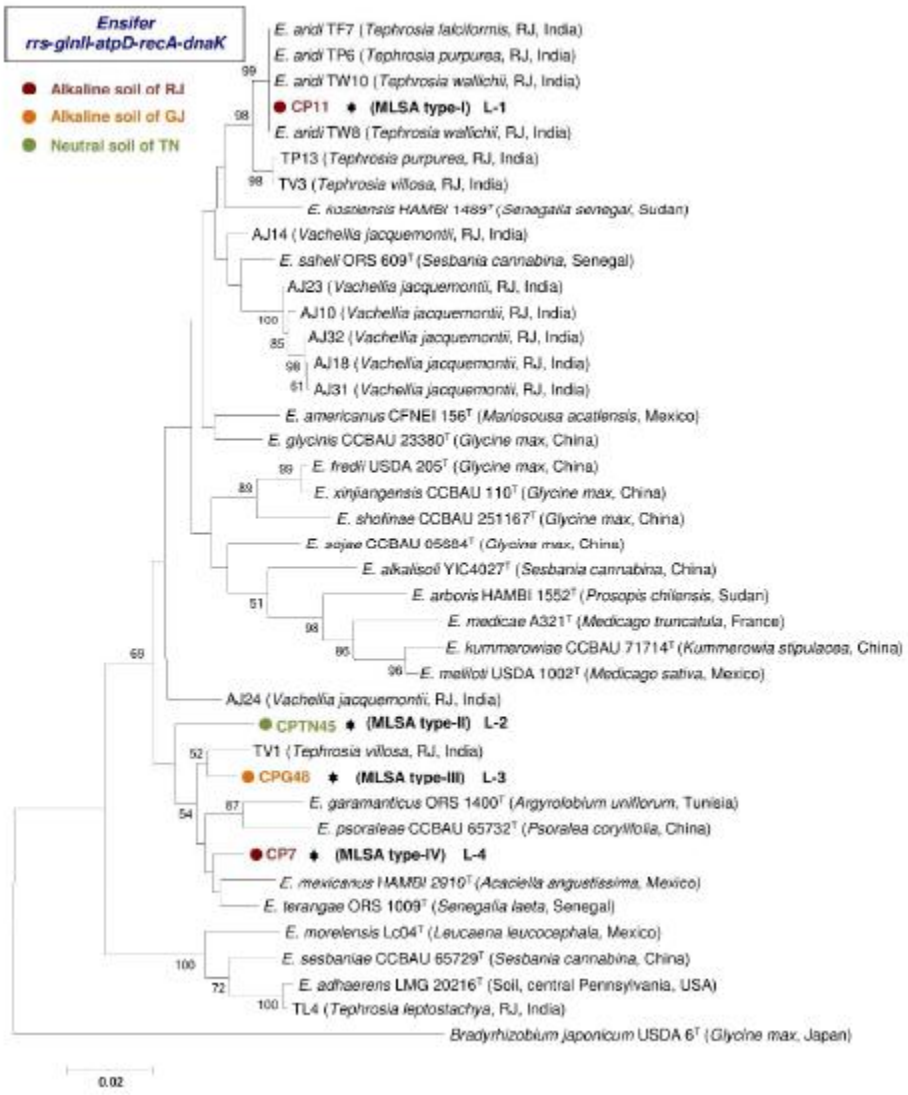

Fig. 3 Maximum Likelihood tree using ***rrs-glnII-atpD-recA-dnaK*** concatenated gene sequences of ***Ensifer*** strains with type strains and close relatives. Bootstrap values more than 50% calculated for 1,000 replications are indicated at internodes. The scale bar indicates 2% nucleotide substitution per site. (Abbreviations: *E.*, *Ensifer*; L, Lineage and ^{T}, type strain)

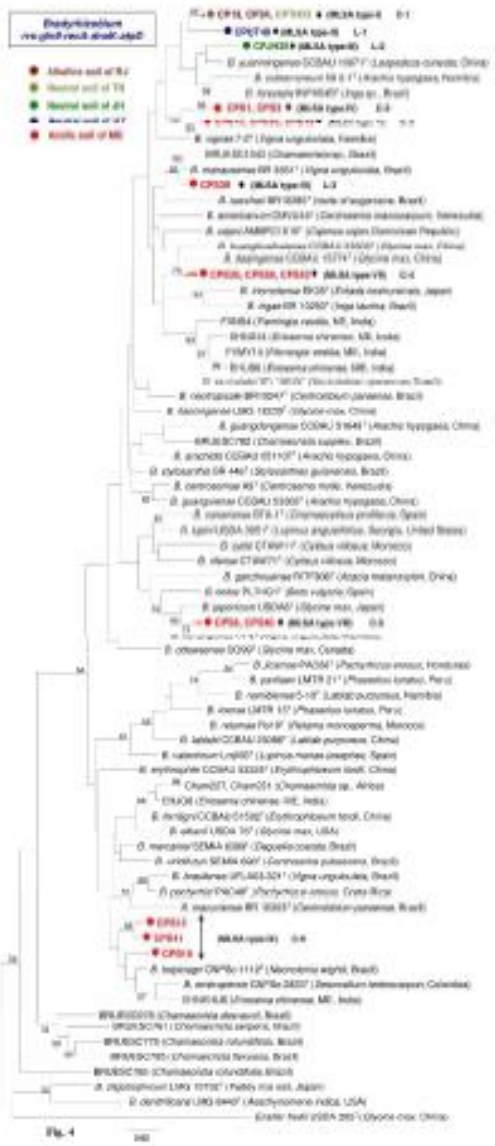

Fig. 4 Maximum Likelihood tree using ***rrs-glnII-recA-dnaK-atpD*** concatenated gene sequences of ***Bradyrhizobium*** strains with type strains and close relatives. Bootstrap values more than 50% calculated for 1,000 replications are indicated at internodes. The scale bar indicates 2% nucleotide substitution per site. (Abbreviations: *B.*, *Bradyrhizobium*; C, Cluster; L, Lineage and ^{T}, type strain)

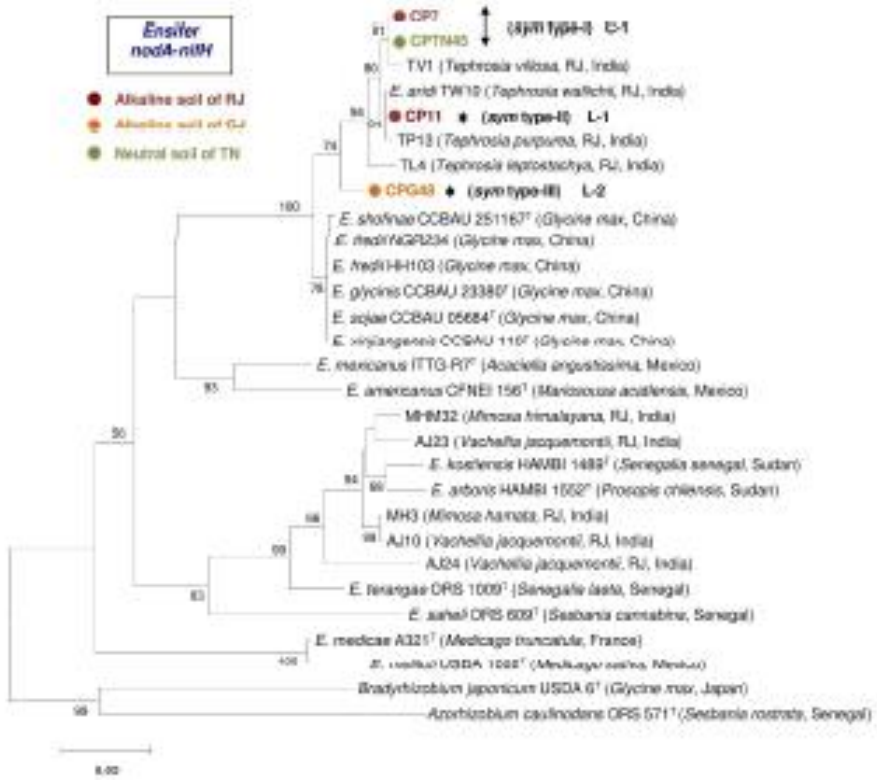

Fig. 5 Maximum Likelihood tree using ***nodA-nifH*** concatenated gene sequences of ***Ensifer*** strains with type strains and close relatives. Bootstrap values more than 50% calculated for 1,000 replications are indicated at internodes. The scale bar indicates 5% nucleotide substitution per site. (Abbreviations: *E.*, *Ensifer*; C, Cluster; L, Lineage and ^{T}, type strain)

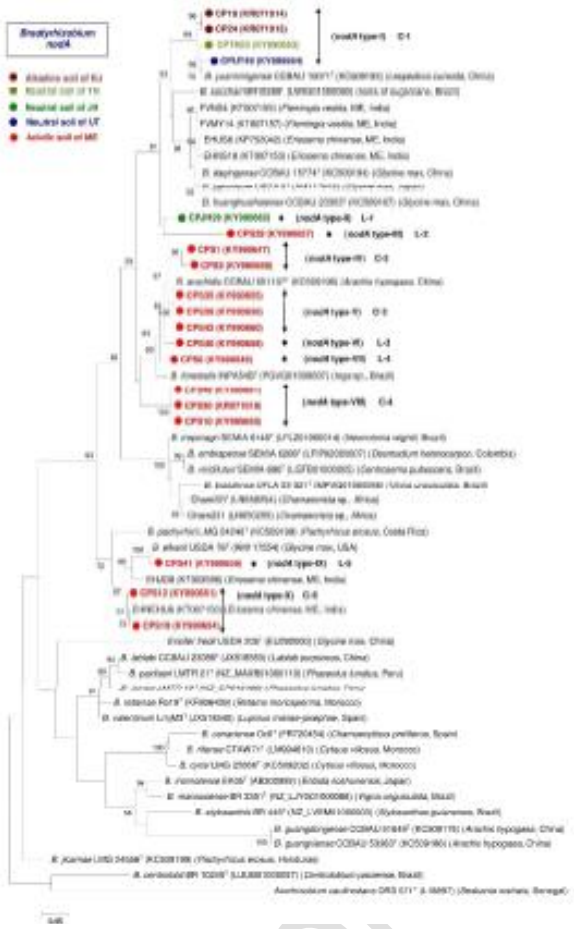

Fig. 6 Maximum Likelihood tree using ***nodA*** gene sequence of ***Bradyrhizobium*** strains with type strains and close relatives. Bootstrap values more than 50% calculated for 1,000 replications are indicated at internodes. The scale bar indicates 5% nucleotide substitution per site. Accession numbers from GenBank are in parenthesis. (Abbreviations: *B.*, *Bradyrhizobium*; C, Cluster; L, Lineage and ^{T}, type strain)

Fig. 7 Maximum Likelihood tree using ***nifH*** gene sequence of ***Bradyrhizobium*** strains with type strains and close relatives. Bootstrap values more than 50% calculated for 1,000 replications are indicated at internodes. The scale bar indicates 5% nucleotide substitution per site. Accession numbers from GenBank are in parenthesis. (Abbreviations: *B.*, *Bradyrhizobium*; C, Cluster; L, Lineage and ^{T}, type strain)

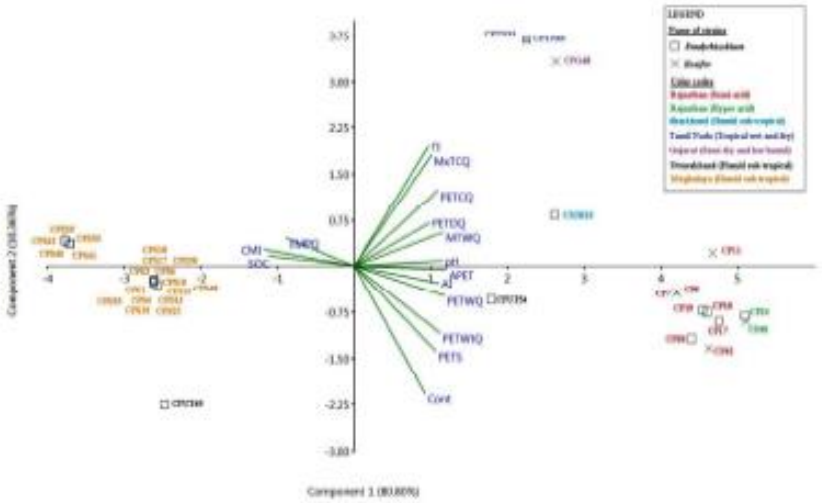

Fig. 8 Biplots of the principal component analysis (PCA) elucidating the correlation of different *Chamaecrista pumila* rhizobial strains with the environmental conditions in different climatic zones of India. *Bradyrhizobium* strains are represented by boxes, while *Ensifer* strains are depicted as cross marks. The strains have been color coded to depict their geographic locations and the climatic zones. Green vector lines along with letters in blue font depict the direction and magnitude of environmental influences on the bacterial strains sourced from different climatic zones in India.

Table 1: Plant nodulation assay results of *Bradyrhizobium* and *Ensifer* strains nodulating *Chamaecrista pumila* in India

Host description			***Bradyrhizobium* strains**										***Ensifer* strains**			
Leguminous plant	**Subfamily**	**Wild (Native/ invasive / introduced) or Crop**	**CP19**	**CP24**	**CPJH29**	**CPTN33**	**CPUT49**	**CPS1**	**CPS6**	**CPS12**	**CPS19**	**CPS30**	**CP7**	**CP11**	**CPG48**	**CPTN45**
Chamaecrista pumila	Caesalpinioideae	Wild (Native)	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+
Vigna radiata	Papilionoideae	Crop	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+
Glycine max	Papilionoideae	Crop	Nod+	Nod+	Nod+	Nod+	Nod+	Nod-	Nod-	Nod-	Nod-	Nod-	NT	Nod-	NT	NT
Cyamopsis tetragonoloba	Papilionoideae	Crop	Nod-	NT	NT	Nod-	NT	Nod-	NT	Nod-	Nod-	NT	Nod+	Nod+	NT	NT
Tephrosia villosa	Papilionoideae	Wild (Native)	NT	Nod+	NT	Nod+	Nod+	NT	NT	NT	Nod+	NT	Nod+	Nod+	NT	Nod+
Tephrosia wallichii	Papilionoideae	Wild (Native)	Nod+	Nod+	NT	Nod+	Nod+	Nod+	Nod+	Nod+	NT	NT	NT	Nod+	Nod+	Nod+
Leucaena leucocephala	Mimosoid clade, Caesalpinioideae	Wild (Invasive)	Nod-	Nod-	NT	NT	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod+	Nod+	NT	NT
Mimosa hamata	Mimosoid clade, Caesalpinioideae	Wild (Native)	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-
Prosopis cineraria	Mimosoid clade, Caesalpinioideae	Wild (Native)	Nod+	Nod+	NT	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+	Nod+
Prosopis juliflora	Mimosoid clade, Caesalpinioideae	Wild (Invasive)	Nod+	Nod+	NT	NT	NT	Nod-	Nod-	Nod-	Nod-	Nod-	NT	Nod+	NT	Nod-
Senegalia senegal	Mimosoid clade, Caesalpinioideae	Wild (Native)	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-
Vachellia jacquemontii	Mimosoid clade, Caesalpinioideae	Wild (Native)	NT	Nod-	NT	Nod-	NT	Nod-	Nod-	Nod-	Nod-	Nod-	NT	Nod-	NT	NT
Vachellia leucophloea	Mimosoid clade, Caesalpinioideae	Wild (Native)	Nod-	Nod-	NT	NT	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	NT	Nod+	NT	NT

Vachellia tortilis	Mimosoid clade, Caesalpinioideae	Wild (Introduced)	NT	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	Nod-	NT	Nod+	NT	NT

Nod+ indicates positive nodulation, Nod- indicates no nodulation and NT indicates not tested